

### 3 The Mechanisms Behind Increased Explicitness in Translations

#### A Multifactorial Corpus Investigation of the *Om*-Alternation in Translated and Original Dutch

*Amélie Van Beveren, Gert De Sutter  
and Timothy Coleman*

#### 1. Introduction<sup>1</sup>

Explicitation is a much-discussed subject in Translation Studies. Since Blum-Kulka's (1986) Explicitation Hypothesis (1986) and Baker's (1993) discussion of explicitation as a translation universal, the relatively larger presence of explicit encodings in translated texts is generally regarded as one of the defining features of translated language. Although this feature can be operationalised in different ways at lexical, syntactic and discourse levels, the present study deals with a case of grammatical explicitness, viz. the use of the optional complementiser *om* in Dutch infinitival clauses.

Baker's seminal paper has triggered a substantial body of empirical research into translational explicitation (see e.g. Olohan 2003 on contractions in translated and non-translated English, Klaudy 2004 on the implied object in English and Hungarian, Pápai 2004 on different types of explicitation strategies such as the addition of conjunctions and discourse particles in English and Hungarian, Hansen-Schirra et al. 2007 on cohesion markers in English and German, and so on). Despite the fact that most of these studies corroborate that explicitation occurs more often in translations than in non-translated texts (also see e.g. Øverås 1998; Olohan & Baker 2000; Kruger 2018), the question as to what *causes* the increased explicitness in translations is still a matter of debate (also see Kruger 2018 for discussion). As a consequence, it is still impossible to predict exactly when explicitation will occur. In a recent study, Kruger & De Sutter (2018), using the variation between English complementation clauses with and without the complementiser *that* in British and South African English as a test case, have addressed the relative contributions of the three most common explanations proposed for the relatively stronger presence of the more explicit option in translated texts: the processing-strain hypothesis (Olohan & Baker 2000), the risk-aversion hypothesis

(Pym 2005, 2015) and the source-language transfer hypothesis (Becher 2010). Kruger & De Sutter (2018) conclude that processing effort and risk avoidance are the most plausible mechanisms behind explicitation of *that* in translated South-African English texts and rule out cross-linguistic interference. More precisely, it seems that translators, in preferring *that* over *zero*, opt for the explicit variant which, in this case, is not only the choice least likely to result in communicative misunderstanding but also the relatively most frequent option *and* the choice most strongly associated with perceived norms of good writing. In the case of the *that*-alternation, the form with *that* can therefore be considered the “safest” option in several respects. In other cases of grammatical variation, though, the most frequent variant is *not* the most explicit one, which thus leads to the question whether translators prefer to explicitly mark the syntactic structure rather than just prefer the most common option. Clearly, a close investigation of such a case of grammatical variation would help to further disentangle the different mechanisms potentially contributing to the larger degree of explicitation typically found in translations.

An alternation that meets this condition is the Dutch *om*-alternation, i.e. the variation between infinitival complements with and without the prepositional complementiser *om*, as exemplified in (1) below:

1. *Hij beslist (om) thuis te blijven* “He decides to stay at home”.

Descriptive and prescriptive grammars state that the implicit form is more common in written language and the explicit form in spoken language (ANS 1997: 1111). In other words, the implicit form is the default option for writers (Vliegen 2001: 127), whereas the explicit form is the safer option in terms of highlighting the syntactic structure of the sentence. A more detailed description of the alternation follows in Section 2.2.

The main goal of this paper is to find out whether risk avoidance affects the choice between implicit and explicit *om* in translated Dutch, and, more particularly, which aspect of the risk-aversion hypothesis is most decisive, i.e. does risk-aversion lead translators to opt for the more explicit variant or for the overall most common variant? Through an application of the Multifactorial Prediction and Deviation Analysis method (henceforth abbreviated as MuPDAR) (Gries & Deshors 2014) to a data set extracted from the Dutch Parallel Corpus (Macken et al. 2011), we will first investigate to what extent the factors underlying the choice between implicit and explicit *om* are similar in translated and non-translated language. Previous empirical analyses with non-translated data have pointed out that the complementiser *om* can be added or omitted depending on different syntactic, semantic and pragmatic factors (Blom 1984; Jansen 1987; Blom 1990; Van Haaften 1991; ANS 1997; Pardoën 1998; Vliegen 2001; Bouma 2017).

We will also try to shed further light on the relationship between the risk-aversion hypothesis and the other proposed hypotheses in the literature,

more particularly by taking into account register and complexity-related factors, and by a brief comparison of our translation data with the equivalents of the *om*-alternation in the two source languages included in the Dutch Parallel Corpus, viz. French and English.

The chapter continues with a section providing further theoretical background, which first gives a brief overview of important existing studies on the mechanisms behind explicitation in translation (Sub-section 2.1) and introduces the Dutch alternation under investigation (Sub-section 2.2). Then, Section 3 outlines the methodology, focusing on the corpus, the compilation of the data set and the quantitative analysis. Section 4 presents and discusses the results of the quantitative analysis, and Section 5 presents the conclusions.

## 2. Theoretical Background

### 2.1. Proposed Causes of Explicitation in Translation Studies

As was already mentioned, existing studies on explicitation in translation differ in opinion about the mechanisms underlying the larger degree of explicitation found in translated language. Kruger (2018) and Kruger & De Sutter (2018) outline three main hypotheses, viz. the psycholinguistic processing-strain hypothesis, the pragmatic risk-aversion hypothesis and the source-language transfer hypothesis. The two studies differ in their methodology: Kruger (2018) tests her data through logistic regression and cluster analysis. Kruger & De Sutter (2018) use MuPDAR, because this statistic technique is specifically aimed at uncovering the variables that cause the divergences between two different language varieties, in their case translated and non-translated English.

The first hypothesis, the processing-strain hypothesis, is the one that is adopted in the classical Olohan & Baker (2000) paper on *that* vs. *zero* complement clauses in translated English. They argue that explicitation occurs unconsciously as a consequence of the psycholinguistically constrained and cognitively demanding environment of translation. In other words, the processing-strain hypothesis is related to cognitive complexity: more demanding cognitive environments, such as translations, are associated with the selection of more explicit and more analytical linguistic options. Kruger (2018) investigates this hypothesis in a data set of *that* vs. *zero* complement clauses in original and translated South-African and British English by investigating the effects of a number of complexity-related factors on the grammatical choice in question, such as the complexity of the matrix-clause subject and of the verb and the distance between the matrix-clause verb and the onset of the complement clause. If the processing-strain hypothesis holds, one expects a stronger effect of complexity factors in translations compared to non-translations. Kruger (2018) finds that such complexity-related factors play a significant role in

the choice between the implicit or explicit option in both translated and non-translated data. According to the processing-strain hypothesis, one should expect to find a relatively larger preference for the explicit option in the translation data in both low and high complexity contexts, as the processing strain associated with the cognitively demanding activity of translation adds to the processing strain related to language-internal complexity factors, and a choice for the more explicit option can be a way of relieving that processing strain (Rohdenburg 1996; Hawkins 2004; Mondorf 2014). Kruger (2018), however, finds that such complexity-related factors play a similar role in both translated and non-translated data. Thus, contrary to what the processing-strain hypothesis predicts, the alternation between *that* and *zero* does not seem to be more sensitive to complexity-related factors in translations compared to non-translations. Rather, translators appear to opt more frequently for the explicit option in general, independent of the complexity level, thereby suggesting a risk avoidance strategy (see the following text).

The second hypothesis is the receiver-oriented risk-aversion hypothesis—in contrast to the processing-strain hypothesis, which is rather producer-oriented. Translation is seen as a high-risk form of communication because the translator functions as a language mediator who has to deal with cultural distance or contexts of fewer shared references between the original writer and the reader of the translated texts. Pym (2005, 2015), Becher (2010), and Kruger & Van Rooy (2016), among others, explicitly link the increased explicitness found in translations to risk-avoidance strategies, i.e. the translator consciously opts for more explicit encodings in translation as a way of reducing the risk of misunderstanding and communicative failure. This hypothesis would imply that translators tend to choose the explicit variant even in contexts in which the implicit variant would otherwise be preferred. Kruger & De Sutter (2018) observe that one way to test the risk-aversion hypothesis is by investigating the influence of register in both translated and non-translated texts. According to the risk-aversion hypothesis, translators will not—or to a lesser extent when compared to non-translators—take into account register differences in the use of the complementiser and choose conservatively for the safest option in order to reduce the risk of misunderstanding. This implies less variation across registers in the presence or absence of the complementiser. In other words, if translated texts show less register sensitivity in comparison with non-translated texts, this could be seen as pointing in the direction of a general preference of translators for the safer option with an explicit complementiser, even in contexts or text types where non-translated writing would typically omit it, with a view toward optimizing the processing ease of the addressee.

A less convincing explanation for the increase of explicitness in translations concerning the *that*-alternation, according to Kruger (2018), is the source-language transfer hypothesis, the third mechanism that is sometimes

proposed in the existing literature. The idea behind this hypothesis is that a source text in a language with a preference for greater explicitness may result in a target text that demonstrates increased explicitness in comparison with non-translated texts in the same language and vice versa (see, e.g. House 2011; Toury 2012; Cappelle & Loock 2013). In Kruger's (2018) study, English and Afrikaans are the languages under investigation. Afrikaans demonstrates a higher omission rate of the complementiser *dat*, the equivalent of *that* in Afrikaans. According to the source-language transfer hypothesis, the expectation would be for translated texts to follow the conventionalised preferences of the source language and thus, in this specific case, for *that*-omission to occur more frequently in English that is translated from Afrikaans in comparison with original, non-translated English. However, Kruger's data refute this hypothesis: English translated from Afrikaans showed a preference for the retention of *that* and, thus, source-language transfer hypothesis was ruled out.<sup>2</sup>

In a nutshell, for many years, three main explanations for increased explicitation in translated texts have been proposed. The multivariate analyses of Kruger (2018) and Kruger & De Sutter (2018) predominantly support the risk-aversion hypothesis and the processing-strain hypothesis: translators are prone to conventionalization or conservatism, meaning that they prefer "safe" discourse options rather than more unusual ones, and more complex grammatical environments are associated with the higher likelihood of explicit constructions.

## 2.2. Om-Alternation

The *om*-alternation refers to the variation between infinitival complements (=IC) with and without the prepositional complementiser *om* in Dutch, illustrated in the examples that follow (see 2), where the infinitival clause depends on a verb, a noun, an adjective or a preposition, respectively. When *om* is present, it functions as a boundary signal that explicitly marks the start of the IC. In some, admittedly fairly rare cases, its presence is crucial to disambiguate the clause structure, as is exemplified in (3). In the absence of an explicit boundary marker, (3a) is ambiguous in that the object *Nora* can either be part of the matrix clause or of the IC; in other words, it is unclear whether Peter promised someone that he would call Nora or that Peter promised Nora that he would call (someone). The variants with *om* in (3b) and (3c) each allow for just one of these two interpretations. In this sense, the IC with *om* can be considered the more grammatically explicit variant, whereas the IC without *om* is the grammatically implicit variant<sup>3</sup> (see e.g. Rohdenburg 1996 for a similar take on grammatical explicitness).

2. a. *Hij probeert (om) een cake te bakken.* "He tries to bake a cake".
- b. *Hij breekt zijn belofte (om) op tijd te komen.* "He reneges his promise to be on time".

- c. *Ik ben bang (om) je kwijt te raken.* “I am afraid to lose you”.
  - d. *Ik slaagde erin (om) het boek uit te lezen.* “I managed to finish the book”.
3. a. *Peter beloofde Nora te bellen.* “Peter promised Nora to call”.
- b. *Peter beloofde om Nora te bellen.* “Peter promised someone that he will call Nora”.
  - c. *Peter beloofde Nora om te bellen.* “Peter promised Nora that he will call someone”.

The *om*-alternation is similar to the *that*-alternation in English in a number of respects: both involve a complement clause that is optionally introduced by a complementiser, the presence or absence of which has no clear semantic consequences. It should be noted, however, that there are also differences. First of all, finite complements (in the *that*-alternation) have an explicit subject, whereas infinitival complements (as in the *om*-alternation) do not. The most important difference between the two alternations is that, in the Dutch *om*-alternation, it is the implicit variant that is the overall most frequent option, and not the explicit variant. While for the *that*-alternation, the explicit variant is associated with informational written language at the formal end of the scale and the implicit option is associated with informal spoken language, the situation for the *om*-alternation is the other way around (Vliegen 2001: 127). Vliegen shows this through a corpus-based investigation of IC's with 11 selected illocutionary verbs, such as *beloven* “promise”, *aanbevelen* “recommend” and *verzoeken* “request” in newspaper data, and finds that in 82.05% of the cases the implicit variant is chosen. The results from Bouma's (2017) corpus-based study corroborate this: in only 23% of the cases *om* is retained in written language (a subset of the Twente Newspaper corpus, consisting of material from *Algemeen Dagblad* and *NRC Handelsblad* from 1994 and 1995). In the corpus used for the present study as well, we found evidence for a general preference for the implicit construction in written Dutch (see Section 3).

As is true for the most (grammatical) alternations, the choice between explicit and implicit *om* is not random: several existing accounts have proposed syntactic, semantic and/or language-external factors that have an influence on the alternation (Jansen 1987; Van Haften 1991; ANS 1997; Vliegen 2001; Bouma 2017). Some examples of syntactic factors mentioned in the literature are the head of the IC and the form and voice of the matrix verb. Semantic factors suggested to play a role include the agentivity of the implied subject in both the matrix clause and the IC, which has a positive influence on the presence of *om* according to Vliegen (2001), and the semantic category of the verb in the matrix clause, with verbs indicating a purpose displaying a relative preference for the explicit construction (Bouma 2017). Register and language variety are potential language external factors: first of all, the implicit construction occurs more often in more formal language than in more informal language

(ANS 1997). Vandeweghe (1972) also mentions a geographical difference: in Belgian Dutch the complementiser is more often omitted when compared to Netherlandic Dutch. Bouma (2017) is the only existing study that investigates the effects of several of such potentially determining factors in a multifactorial corpus-based investigation of the absence or presence of *om* in present-day written language. This approach is well suited because explicitation is not a one-dimensional phenomenon, but rather depends on a number of variables (Serbina 2015: 37). Bouma concludes that the combination of a set of complexity-related factors such as the number of intervening words between the head of the IC and the onset of the IC and a number of semantic factors related to the governing verb produces the best model for predicting the variation between the explicit and implicit construction.

In this paper, we mainly focus on the factors that can help us disentangle the three main causes for explicitation in translations. More details about the variables relevant for our analysis are provided in Section 3.3.

### 3. Methodology

#### 3.1. *Dutch Parallel Corpus*

The corpus used for this case study is the Dutch Parallel Corpus (henceforth abbreviated as DPC), which contains approximately ten million words of running text and is a multifunctional, multiregister and bidirectional parallel corpus, consisting of both translations into Dutch in several registers and comparable texts originally written in Dutch (Macken et al. 2011). The texts are contemporary and include translations from English and French. Originally, the corpus contained five different text types: literature, journalistic texts, instructive texts, administrative texts and external communication. Following Biber & Conrad's methodology (2009: 40), however, Delaere (2015) reorganised these registers, by selecting four situational and non-linguistic characteristics: addressor, addressee, channel and communicative purpose. On the basis of these characteristics, the texts were redistributed in the following eight different registers: specialised communication, broad commercial texts, fiction, journalistic texts, instructive texts, political speeches, legal texts and tourist information. For this case study, we did not include the registers specialised communication and broad commercial texts, because both consist of a rather broad range of text types which it is difficult to posit generalizations about. In Table 3.1, we present an overview of the six remaining registers that were included in the present study, with some text type examples.

#### 3.2. *Data Extraction*

The DPC can be exploited either as a full text resource or via a web search interface.<sup>4</sup> The corpus cannot only be consulted by searching for

Table 3.1 Overview of the different registers represented in the sample

Register	Text type examples
Fiction	Novels
Journalistic texts	Comment articles (columns) Essayistic texts News articles
Instructive texts	Manuals DIY guides
Legal texts	Internal legal documents Legislation Descriptions of legal procedures
Political speeches	Official speeches Proceedings of parliamentary debates
Tourist information	Informative documents of a general nature

Table 3.2 Search query in DPC for the *om*-alternation

Wildcard	Part 1	Wildcard	Part 2	Wildcard	Part 3	Wildcard
/	V, N, Adj	0–10	<i>i) te</i>	0–2	V, flag: inf	/

specific word forms or lemmas, but one can also do queries based on the morphosyntactic class of a given word. Because we distinguish four different types of the constructions under investigation according to the head of the IC, we queried the DPC via a word class-based search as the first part of a multiple search entry. We selected the word classes of *verb*, *noun* and *adjective*. *Preposition* was left out because relevant hits with a preposition as the head of the IC will be included in the results for the three other word classes (as the preposition itself necessarily depends on a verb, an adjective or a noun). The verb, noun or adjective must be followed by *te*, which is a fixed element of the Dutch IC, within a range of ten words. The query also stipulates that *te* must be followed by an infinitive within a range of two words. Ranges are added in the form of a wildcard. A wildcard also needs to be placed before and after the search string, in order to allow random words to precede or follow the string. In Table 3.2, a schematic overview of the query is given.

Such a general search query inevitably generates a great deal of noise that requires intensive manual post-processing, but it guarantees a good recall. The search query returned a total of 40,100 hits. The majority of these had to be removed and ca. 12,500 relevant instances of IC's with or without the complementiser *om* remained. This means that the construction occurs with a normalised frequency of 989.22 relevant constructions per one million words of running text for non-translated data and 1,172.46 relevant constructions per million words for translated data. The hits returned by the query were considered relevant when (1) they consist of an IC introduced by *om* in which *om* could be omitted without



a change in meaning or grammaticality, or, (2) the other way around, of an IC without a complementiser where *om* could be added without producing a change in meaning or in grammaticality. In case of doubt about the possibility of adding or deleting *om*—for instance, when, for a certain head, only cases with an implicit IC are present in the corpus data but the explicit form is allowed according to our linguistic feeling—we searched for evidence via Google. This was the case for verbs such as *hopen* “hope” and *wensen* “wish”: in the corpus, these verbs are only combined with implicit IC’s, but via Google queries, we found several instances of explicit IC’s governed by the verbs in question, testifying to the possibility of this option in present-day Dutch.

As a full annotation of all ca. 12,500 remaining instances for a number of different semantic and formal factors (See sub-section 3.3) would have been infeasible, we extracted a random sample of 1,734 sentences for this study, which was manipulated to make sure that enough attestations of each register under investigation are represented: more precisely, the sample was composed so that each of the six selected registers is represented by about 300 relevant observations of the *om*-alternation. The distribution of the construction over the selected registers and over translated vs. original Dutch in our sample is given in Figure 3.1.

The distribution as visualised previously will guarantee a register-wise comparability of the original and translated parts of the sample, which contain 745 and 989 hits, respectively. The statistics of the distribution of the sample will be discussed in Sub-section 4.1.

### 3.3. Selected Variables

Both samples—one with the translated data and one with the non-translated data—were subsequently annotated for a set of linguistic variables. The

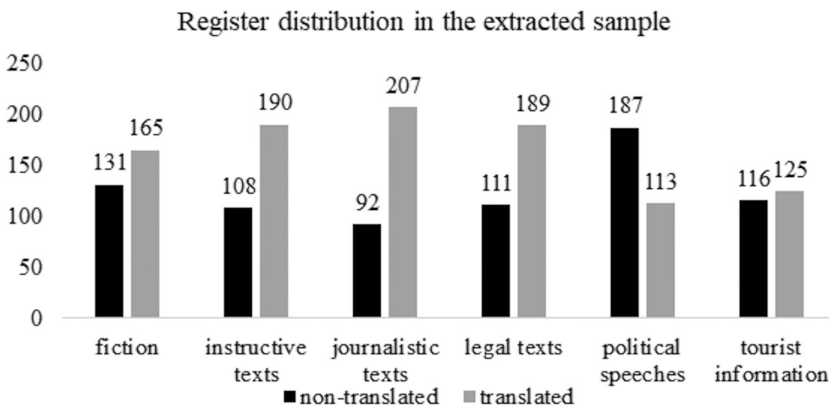


Figure 3.1 Register distribution according to the translation status in the extracted sample

dependent variable in our study is the binary choice between retention versus omission of the complementiser *om*. The variables or predictors included in the investigation were selected on the basis of existing literature on the *om*/zero alternation and/or on explorative corpus investigations, and with an eye on the hypotheses under scrutiny. In Table 3.3, an overview of the selected variables is given.

Most of the language-internal factors identified as determinants of the variation between retention and omission of *om* in the existing literature revolve around the processing load caused by the structure of the matrix clause and the IC: the level of embedding of the IC (i.e. is it directly subordinate to a main clause or is it more deeply embedded), the form and voice of the verb of the matrix clause, whether or not the matrix verb is a separable compound verb, the distance between the IC and the matrix clause head on which it is dependent and the length of the IC. According to Rohdenburg's (1996) complexity principle, the more explicit option is generally preferred in cognitively more complex environments, i.e. respectively, in more deeply embedded IC's, in IC's depending on a passive rather than active matrix verb, in IC's depending on an infinite rather than finite matrix verb, in IC's depending on a separable complex verb and in contexts with a large number of intervening words between the IC and its head or with a lengthy IC. We also added a coordination variable, viz. whether or not the IC is the second, third, etc., so the non-first, in a series of coordinated IC's. This variable is not to be taken as just another complexity-related factor. While the presence of more IC's in a single sentence is an indication of increased complexity, we do not automatically expect more explicit constructions in these environments, because there

Table 3.3 An overview of the selected variables and their values

Variable	Values
Coordination: IC is the 2nd, 3th, etc. in a series of coordinated IC's	Yes, no
Level of embedding: IC is directly subordinate to a main clause	Yes, no
Form of the matrix verb	Finite, infinite
Voice of the matrix verb	Active, passive
Separable compound matrix verb	Separated, not separated
Distance between the head of the matrix clause and the IC (in number of words)	0, 1, 2, 3 . . .
Syntactic complexity of the IC (in number of words)	0, 1, 2, 3 . . .
Register	Fiction, journalistic texts, legal texts, instructive texts, political speeches, tourist information
Source language	French, English
Head IC	V, N, Adj, Prep

is a tendency to avoid repetition of the complementiser in coordinated contexts.<sup>5</sup> In other words, we expect to find relatively fewer instances of explicit *om* in such a series of coordinated IC's.

Like Kruger (2018) and Kruger & De Sutter (2018), we also added register to our model. Register as it happens has repeatedly been shown to influence the likelihood of complementiser omission (De Sutter et al. 2005; Tagliamonte & Smith 2005; Torres Cacoullos & Walker 2009; Biber 2012; Kolbe-Hanna & Szmrecsanyi 2015). On the basis of the different effects of the register factor in translated and non-translated data, Kruger (2018) and Kruger & De Sutter (2018) conclude that the risk-aversion hypothesis is the most convincing mechanism behind increased explicitness in case of the *that*-alternation because translators can be seen to display less register-sensitivity (See sub-section 2.1).

We obviously also take into account source language as one of the possibly determining factors. As already mentioned, Kruger & De Sutter (2018) found no evidence for the source-language transfer hypothesis in their data on *that/zero* in English translated from Afrikaans. Since we deal with another phenomenon and other language pairs in this study, viz. Dutch-French and Dutch-English, we cannot ignore the possibility of source language influence. We will conduct a preliminary analysis of the English and French source text equivalents of the *om/zero* attestations in the translated Dutch corpus attestations and analyze the influence of source language on the choice between the explicit or implicit option in the translated data set.

Finally, we take into account the variable “head of the IC”, as this is generally considered a crucial variable in the modelling of the *om/zero*-variation. Bouma (2017), for instance, clearly shows that this factor plays a significant role in the choice between the explicit or implicit variant in original Dutch. The present study does not yet include semantic factors such as the presence of evaluating adverbs like *helaas* “unfortunately” and *gelukkig* “happily” and the semantic category of the IC's head, though Bouma's (2017) study has indicated that these potentially have a role to play as well. We aim to include these in future research.

### 3.4. *MuPDAR Analysis*

Like Kruger & De Sutter (2018), we opt for the Multifactorial Prediction and Deviation Analysis (henceforth abbreviated as MuPDAR) method (Gries & Deshors 2014). As Kruger & De Sutter (2018: 251) state, MuPDAR “represents an influential methodological advance in studying variation in contexts where linguistic choices in a ‘peripheral’ variety are directly studied in relation to the ‘central’ variety”. Kruger & De Sutter (2018: 261) argue that this method can be successfully extended to the context of Translation Studies with translated language as the more peripheral variety and the original language as the central one. The software used for the statistical analyses to be presented below is R 3.5.1.

In the MuPDAR procedure, two different generalised linear mixed-effects models (abbreviated as glmm) are fitted. The first glmm is based on the non-translated data only; its output gives insight into the variables that significantly influence the choice between the explicit and implicit construction in non-translated data. In a next step, we measure how well this model can predict the choices made in the translated data. In the third and final step, a second glmm is fitted, this time on the translated data, with the same fixed effects as the first glmm and a new response variable, viz. the correctness of the predicted choices in translated texts on the basis of the model of the non-translated data set.

One of the practical limitations of the MuPDAR approach is that only factors that possibly play a role in both data sets—i.e. in this case, translated and non-translated Dutch—can be added. In other words, source language, a factor that is only relevant for translated language cannot be included. For that reason, we will fit a third glmm based on the translated data only, in which we will take into account the source language variable next to all the other factors under investigation as listed in Section 3.3 above.

For all statistical tests performed in this study, the significance cut-off level is set at 0.05. In other words, all p-values smaller than 0.05 indicate statistical significance, p-values larger than 0.05 indicate non-significance. In the analysis, we use a logarithmic transformation for the two numeric variables, to reduce skewing and outliers in the data, as is customary in quantitative modelling (Baayen 2008: 31).

In the remainder of this article, the focus is not on the technical details, but on the interpretation of the analyses. For further information about the technique behind the MuPDAR-analysis, see Gries & Deshors (2014), Gries & Bernaisch (2016) and Kruger & De Sutter (2018).

## 4. Results

In this section, we first focus on the distribution of the explicit and implicit constructions in both translated and non-translated Dutch in order to find out whether both varieties significantly differ from each other and in which direction. This will enable us to formulate an answer to the question whether translators indeed show a relatively greater preference for the explicit variant (Sub-section 4.1). The remainder of the section is structured as follows: we first present a glmm of the non-translated data, which we assume to be the central variety (Sub-section 4.2). This model is then used as the baseline for the next stages of the analysis: in Sub-section 4.3, we first investigate how successful the model built on the basis of the non-translated data is in predicting the choices of the translators in the translation data and then, on the basis of the best fitted model of the non-translated data, we fit a new glmm that analyses which predictors contribute to the deviance of the translated data from the non-translated data. These predictors can help to elucidate

the reasons behind the different linguistic behavior of translators and non-translators. The results of the MuPDAR analysis can shed important light on the relevance of the risk-aversion hypothesis and the processing-strain hypothesis for the case of grammatical variation under discussion here. In Sub-section 4.4, we turn to a discussion of the source-language transfer hypothesis.

#### 4.1. *Translated vs. Non-Translated Dutch*

Before we present the results of the MuPDAR analysis in the next subsection, we first test whether the translated and non-translated data are significantly different from each other in their preferences for the explicit or implicit construction. Figure 3.2 presents the distribution of the sample.

First of all, the difference between the translated and non-translated data set in their relative preferences for the explicit or implicit construction is highly significant (X-squared = 13.18,  $df = 1$ ,  $p\text{-value} = 0.000283$ ). This overall frequency contrast corroborates that Dutch translators use the complementiser *om* significantly more frequently than non-translators do. In non-translated Dutch the complementiser is omitted in 56.64% of the cases, while in translated Dutch, this is only the case in 47.72% of the relevant instances. Thus, translators are shown to prefer, at least in the case of the Dutch *om*-alternation, the most explicit option rather than the default option (when considering the distribution in non-translated Dutch as the baseline). This overall difference in frequency ties in with the processing-strain hypothesis (i.e. translators reduce the processing effort for themselves by using an explicit boundary marker) as well as the

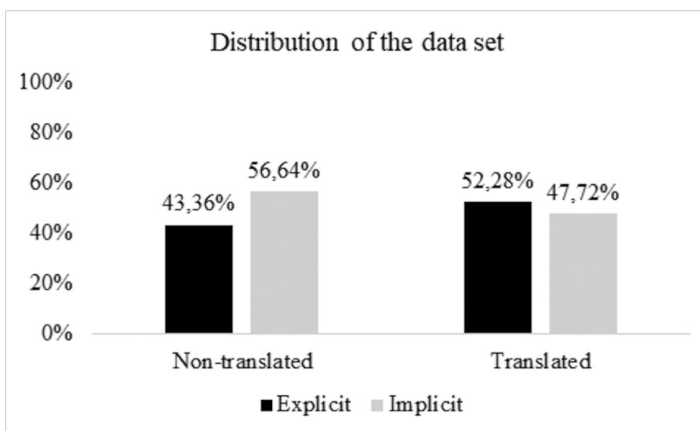


Figure 3.2 Distribution of the data set (relative frequencies based on the investigated sample)

risk-aversion hypothesis (translators use the explicit variant in order to reduce the risk of misinterpretation) and hence cannot in itself be used to disentangle these explanations.

#### 4.2. Significant Variables and Effect Directions in Non-Translated Dutch

As already mentioned in Sub-section 3.4, the first step in the comparison between translated and non-translated data is fitting a glmm on the non-translated data that is to serve as the baseline of the MuPDAR analysis. The selected fixed effects are the head of the IC, register and the complexity-related factors such as the distance between the head of the IC and the onset of the IC outlined in Sub-section 3.3. As random effects, we selected the author/text provider and the lemma of the head on which the IC depends.

The analysis shows that only three out of the nine selected predictors—we do not take into account source language in this part of the study—have a significant effect on the choice between the explicit and implicit construction in the non-translated Dutch data. Nevertheless, these three predictors are able to describe and predict the variation in the data set to a large extent: the descriptive accuracy of this model is 82.6%, and its predictive ability is very high ( $c = 0.91$ ). The difference between the fitted model and the intercept-only model, i.e. the model without predictors that simply predicts the most frequent value of the response variable in all cases, is highly significant ( $p\text{-value} < 0.001$ ).

Figures 3.3 to 3.5 show the directions and sizes of these three significant effects.

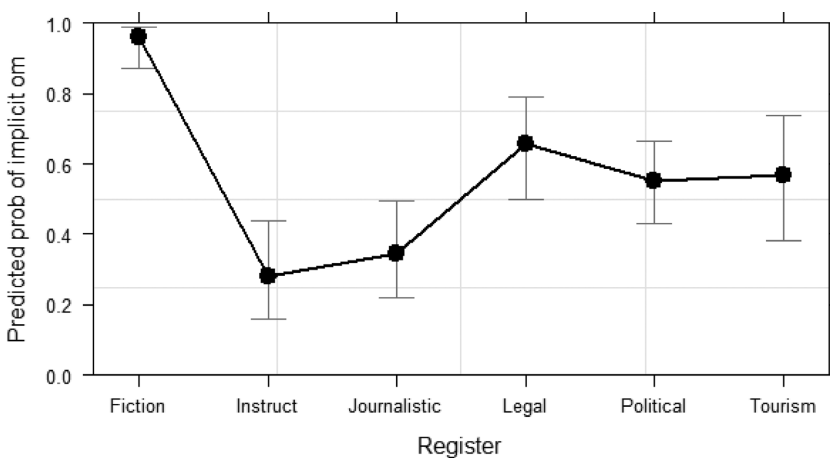


Figure 3.3 Effect plot of the significant predictor Register in non-translated Dutch

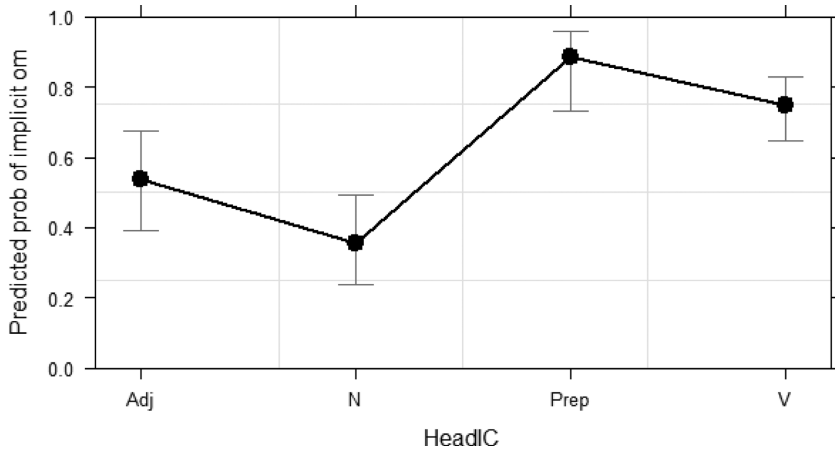


Figure 3.4 Effect plot of the significant predictor Head of the IC in non-translated Dutch

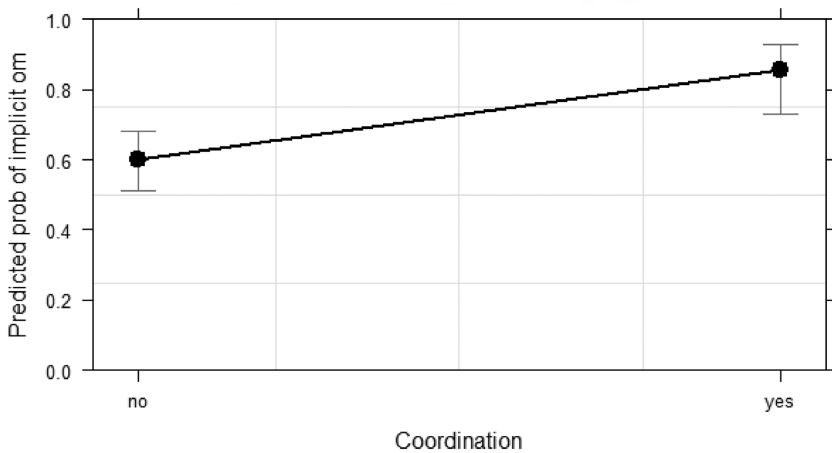


Figure 3.5 Effect plot of the significant predictor Coordination in non-translated Dutch

The effect plot in Figure 3.3 shows that the preference for the explicit or implicit construction depends on the register (as has also been shown to be the case in all kinds of other grammatical alternations, see e.g., De Sutter et al. 2005; Tagliamonte & Smith 2005; Torres Cacoullous & Walker 2009; Biber 2012; Kolbe-Hanna & Szmrecsanyi 2015). The probability of the implicit construction increases most strongly in fiction. Tourist information, legal texts and political speeches also display a general preference

for the implicit construction, but much less outspoken than in the case of fiction. In journalistic and instructive texts, by contrast, the explicit option is favored. However, given that the confidence intervals of all registers but fiction overlap, we can conclude that what characterises the register effect is predominantly the difference between fiction and the other registers. Within that second group of registers we can make a distinction between journalistic texts and instructive texts on the one hand and legal texts, political speeches and tourist information on the other hand. Journalistic texts and instructive texts are the only registers with a preference for the explicit construction and are thus supposed to share certain situational characteristics that the other registers included in the investigation do not share, and that seem to produce a certain tendency toward explicitation in their writers. It remains to be seen what exactly those crucial characteristics are; the same applies to the characteristics setting apart fiction from the other registers. The most important observation, for now, is that register comes out as a significant predictor in the non-translated data set.

In Figure 3.4, the effect plot of the second significant factor, the head of the IC is given. The results are completely in line with the findings of Bouma (2017). IC's with a preposition or a verb as their head have a clear preference for the implicit construction. When the head of the IC is a noun, by contrast, there is a higher chance that the language user opts for the explicit construction. IC's with an adjectival head, finally, display a small preference for the implicit construction.

The final predictor that turned out to be significant is the presence of a coordinated IC. The effect plot demonstrates that the presence of a second IC that is coordinated to a first IC in the same sentence increases the probability of omitting *om* in the second IC. As was already explained, this is not a complexity-related factor on a par with the other complexity-related factors included in the analysis: there are specific reasons why, in this specific syntactic context, language users tend to omit *om*.

In sum, language users appear to decide on *zero* or *om* based on just three of the investigated factors: the register, the head of the IC and the presence of absence of one of more coordinated IC('s). Claims about the latter variable only count for the not-first IC of a series of coordinated IC's. Remarkably, especially against the background of what Kruger (2018) and Kruger & De Sutter (2018) found for the *that*-alternation and what Bouma (2017) showed for the *om*-alternation,<sup>6</sup> none of the "genuine" complexity-related factors seems to play a role. The alternation here is not dependent on syntactic complexity, as operationalised by means of the length of the IC and the distance between the head of the matrix clause and the beginning of the IC. This might suggest that language producers do not use the explicit form to ease cognitive processing for themselves; rather the choice is dependent on stylistic factors and on other morphosyntactic properties, not directly related to syntactic complexity, as discussed prior.



### 4.3. Differences Between Translated and Non-Translated Dutch

In this section, we apply the fitted model of the non-translated data to the translated data set to find out to what extent the model of the non-translated data set can successfully predict the translator's choices. Additionally, it is investigated which factors contribute to the difference between the translated and non-translated data, and in which ways. This second step in the MuPDAR analysis points out that the model built on the basis of the non-translated data is able to predict a mere 59.3% of the choices made in the translated data set correctly. In other words, in 40.7% of the cases, the translator makes another choice between the implicit or explicit construction than a non-translator would have been predicted to do under the same circumstances. The next step in the MuPDAR procedure is to fit a second regression model in order to scrutinise which factors cause that difference. In this second regression model, the response variable is the correct vs. incorrect prediction; the fixed and random effects entered in the model remain the same. Figure 3.6 shows the effect plots of the predictors that contribute to the difference between translated and non-translated texts in a significant way with on the Y-axis the predicted probability of deviant choices (with 0 indicating no deviance compared to non-translated Dutch).

The two predictors that cause translators to make other choices are both complexity-related: the voice of the verb in the matrix clause and the number of intervening words between the head of the IC and the onset of the IC itself, two factors that did not turn out to be significant predictors in the model for non-translated data. The effect plot for voice shows a deviation in both active—associated with a low level of complexity—and passive contexts—associated with a higher complexity. However, the deviance is larger when the verb of the matrix clause is active. This means that translators' choices deviate more often from

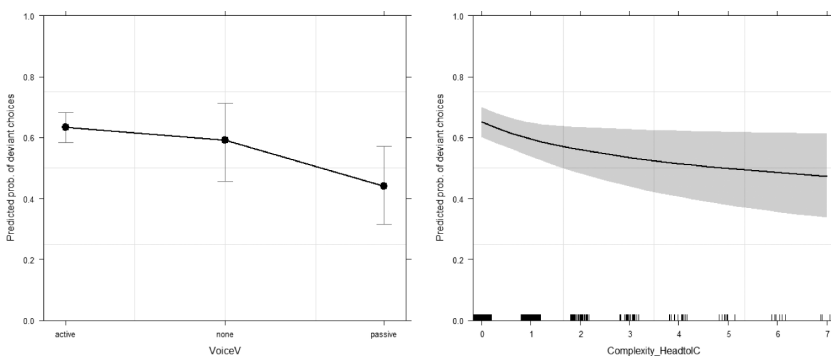


Figure 3.6 Effect plots of the predictors in which translated and non-translated Dutch significantly differ from each other

those of non-translators in active than in passive contexts. The direction of those deviations will be discussed next. The effect plot of the second significant predictor, viz. the number of intervening words between the head of the IC and the IC, shows that translators' choices deviate most from those of non-translators in *low*-complexity contexts, i.e. when the distance between the head of the matrix clause and the onset of the IC is small. Both deviation plots show, in other words, the same tendency: translators deviate more often from non-translators in low complexity contexts. In order to better understand why this is the case, we created deviation plots for both factors, see Figure 3.7.

The plots in Figure 3.7 visualise for each factor level the direction of the deviating preference, with zero signifying no deviation. The deviation score quantifies the size and direction of incorrect predictions. More precisely, if the deviation score is higher than zero, the translator's choice deviates from the non-translators' choices in that he or she opts for the implicit construction in contexts where the non-translator would have been predicted to use the explicit construction; if the deviation score is lower than zero, this means that the translators prefer the explicit construction in contexts where in the original texts the implicit construction would have been preferred. The fact that more observations are attested in the lower half of the left effect plot indicates that in active contexts, the translators significantly more often choose the explicit construction. The same is true when no matrix verb is present in the data. In the case of a passive matrix verb, conversely, the translators relatively more often omit *om*, although the deviations are relatively smaller in comparison with the deviations in the direction of the explicit variant in active contexts.

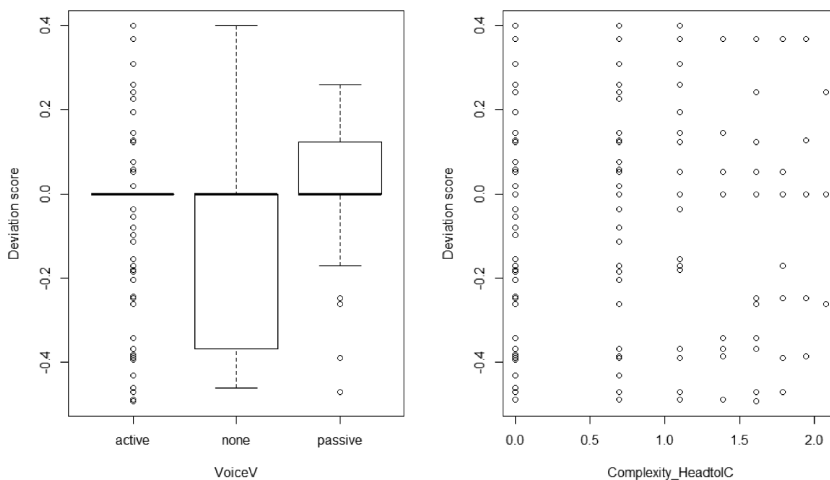


Figure 3.7 Frequency plot of the deviation scores from the glmm for Voice and Complexity\_HeadtoIC

In the right plot, as well, there are more attestations in the lower half of the plot, especially in the left-most column, which includes the cases with distance zero between the head of the matrix clause and the onset of the IC. This again shows that especially in contexts of relatively low complexity, the translators more often opt for the explicit variant than the non-translators. If, as is assumed in the producer-oriented processing-strain hypothesis, translators generally display a preference for the more explicit option in order to reduce processing effort for themselves, we would have expected to see at least a similarly high deviation in the direction of a preference for the explicit construction in contexts of high complexity as well. The idea behind this expectation is, as already mentioned, that translators experience a larger cognitive effort compared to non-translators due to the bilingual language processing characteristic of translating and that the selection of the explicit variant can be a way of reducing that larger processing effort in all contexts, especially in high-complexity contexts. As this is not the case, however, the processing-strain hypothesis must be refuted, at least for the time being. Rather, the results of our analysis could be taken as evidence for the risk-aversion hypothesis, since translators appear to opt in most contexts, independent of the level of complexity, for the most explicit variant; the somewhat higher preference for *om* omission in passive clauses is a notable exception.

Remarkably and in contrast with the results of Kruger (2018) and Kruger & De Sutter (2018), register does not have a significant effect on the amount of the deviation between translated and non-translated Dutch, which means that generally the same register preferences for the implicit or explicit construction could be found in both data sets. According to the risk-aversion hypothesis, we would have expected translations to demonstrate generally less sensitivity to register. However, as we already mentioned in the description of the methodology, the effect of register is only one possible way to investigate the influence of the risk-aversion hypothesis. The most important rationale of this hypothesis is that translators opt for the explicit variant in contexts in which a non-translator would not do that, because the explicit variant is seen as the construction that reduces the risk of misunderstanding and thus helps the receiver to process the sentence.

#### 4.4. Possible Source Language Effects

As already indicated, the MuPDAR analysis cannot test the possible influence of source language transfer, since it does not allow for the inclusion of factors that can only play a role in one of the two data sets under comparison. However, a chi-square test shows that translations from English and translations from French differ significantly in their preferences for the explicit or implicit construction ( $X^2 = 9.17$ ,  $df = 1$ ,  $p\text{-value} = 0.00246$ ). As Figure 3.8 shows, translators from French more often retain *om*, while

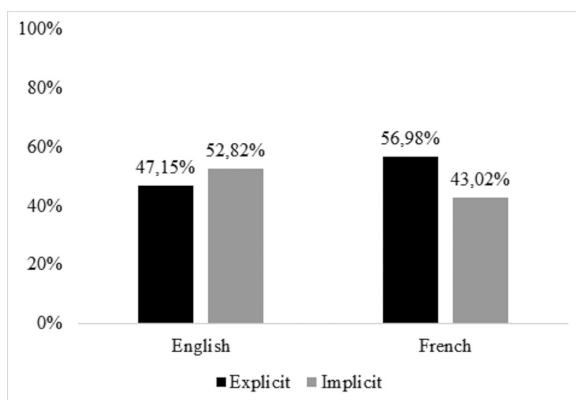


Figure 3.8 Distribution of Source Language in the translated sample

translators from English in about 53% of the cases opt to leave out the complementiser.

We fitted a third glmm for this sub-section based on the translated data only. Its output gives insight into the variables that significantly influence the choice between the explicit and implicit construction in translated data. Factors that turned out to be significant are the head of the IC, coordination, the form of the matrix verb, register and, for now most importantly, source language. In Figure 3.9, the effect plot of source language is shown.

Figure 3.9 corroborates that translators from English tend to omit the complementiser more frequently than translators from French, even if the other factors are controlled for. With this result in mind, it is useful to consider the source language equivalents of Dutch *om*. At first sight, in both English and French, the default option in the infinitival clause is an explicit construction, if this is defined as a construction in which there is a free morpheme present between the matrix clause and the IC. However, in French, the situation is clearer than in English. In the French source data, almost all cases in which an IC is present in the French original contain an explicit boundary signal between the head of the IC and the IC, as illustrated in the following examples:

4. Les parents ont tendance à dire aux enfants qu'ils travaillent afin de gagner de l'argent pour les élever.  
(Eng.: Parents tend to tell their children that they work to earn money for their education.)
5. Il est donc important d'intégrer la gestion du projet dans le processus d'évaluation de ceux qui l'exécutent, afin de les responsabiliser dès le départ.

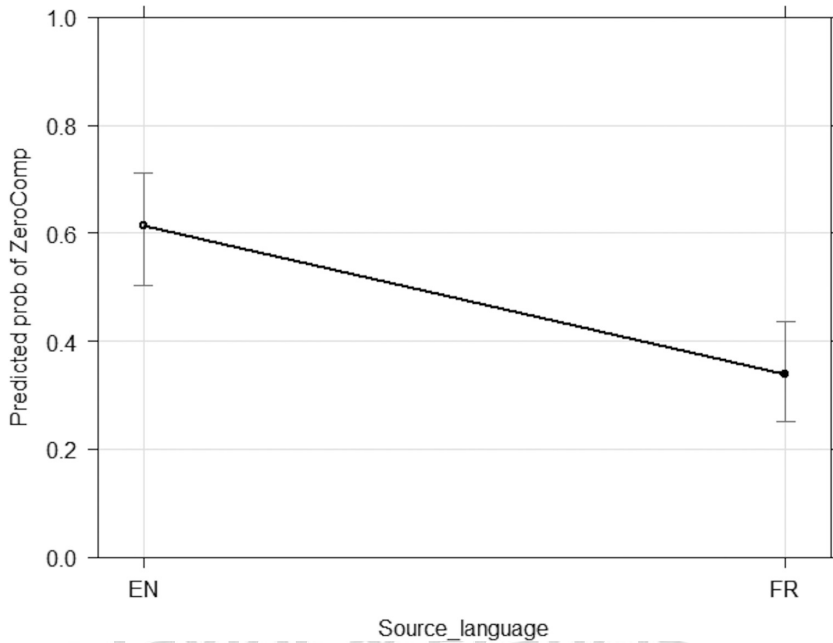


Figure 3.9 Effect plot of Source Language in translated Dutch

(Eng.: It is important to include project management in the evaluation process of involved people in order to empower them from the beginning.)

6. Ils peuvent aussi se mettre d'accord **pour** désigner un expert (géomètre-expert immobilier, architecte, . . .).

(Eng: They are allowed to decide together to appeal to an expert [land surveyor in estates, architect, . . .].)

The French situation is rather different from the Dutch pattern: in French, few verbs allow an optional complementiser. Grevisse & Goosse (2016: 1118–1122) present separate lists of verbs that allow *à*, *de* or *zero* to precede an IC. The lists of verbs *without* a complementiser is remarkably smaller than the lists of verbs that are obligatorily combined with a complementiser. In a mere handful of cases only, the complementiser is *optionally* present, usually with an associated difference in style. For example, *aimer* “to love” can occur with or without the complementiser *à*. Cox (1983: 172) states that infinitival complements without a complementiser most commonly occur in sentences with so-called *verbes du type auxiliaire*, *verbes de déplacement* and *verbes de perception*. In our data,

only in a rare amount of cases (4 of the 222 relevant French source sentences in this case study), no explicit boundary signal is present:

7. Bien sûr, on ne peut pas connaître tous les journalistes, et c'est pour-quoi il **vaut mieux** concentrer ses efforts sur quelques personnes, plutôt que de vouloir à tout prix toucher la majorité.

(Eng.: That is why it is better to concentrate your efforts on some people instead of to keep in touch with everybody.)

8. Qu'**espérez-vous réaliser** aujourd'hui, dans le mois ou l'année qui viennent?

(Eng.: What do you hope to realise in one day—or one month or one year?)

In English, the situation is more complicated. The most frequent equivalent of both the implicit and explicit *om*-construction is a *to*-infinitival clause in English. *To* could be seen an explicit free morpheme between the head of the IC and the onset of the IC. English source sentences *without to* or another boundary signal appear in a few cases only (12 of the 473 English source sentences in this study). According to Huddleston & Pullum (2006: 213), bare infinitival clauses occur in only a very limited set of functions and with a limited set of verbs like *to help*. In (9) and (10), an example of a construction with and without *to* in English is given, respectively.

9. "This study **aims to** discover who is especially good and bad at this mysterious skill, as well as providing some insight into how people can improve their intuitive abilities."
10. "My site **helps** me understand people's feelings, and I wouldn't be without it," Cleave says.

However, *to* can hardly be considered an exact equivalent of Dutch *om* and in a sense is perhaps better regarded as the English equivalent of the Dutch infinitival marker *te*, which, unlike *om*, is obligatorily present in all Dutch IC's. The contrastive grammar by Devos et al. (1992: 220), for instance, explicitly treats *te* as the exact equivalent of English *to*. In English, the status of *to* as a complementiser is a much more discussed subject: it has been treated as a preposition, an infinitive marker/particle, a subordinator or a complementiser (Pullum 1982; Falk 2000; Stevenson 2010). The words *à*, *de* and *pour* are definitely considered to be complementisers in French (Cox 1983: 175) and can be more straightforwardly seen as explicit boundary signals on a par with Dutch *om*. We added the ambiguity of the complementiser as a new factor to the glmm with a test sample of 300 translated sentences from French and 300 translated sentences from English. We annotated the data for three factor levels:

*to*-infinitives are automatically labeled as “ambiguous” for the reason given; other explicit constructions with “compl\_inf”;<sup>7</sup> implicit source texts receive the label “no\_compl”. The factor turned out to be significant as visualised in Figure 3.10.

Source texts with ambiguous *to* are more frequently translated with an implicit construction. In the other cases, the translator more often opts for the explicit option—even when no complementiser is present in the source text—as already mentioned, such sentences barely occur in our data set (20 of the 600 investigated sentences of which the translation result in 9 explicit constructions and 11 implicit constructions). However, further research on the causes behind that difference between sentences translated from English and French is definitely needed.

The results in the present study differ from Kruger’s (2018) in her study of the *that*-alternation in English and Afrikaans. Because Afrikaans is a language with a clear preference for the implicit construction (see Footnote 2, however), she tested whether, as would be predicted by the source-language transfer hypothesis, there was a similar preference for the implicit option in English translated from Afrikaans. However, she found no such effect: English that is translated from Afrikaans has an overall preference for the explicit construction, which, following Kruger & De Sutter (2018) can be seen as a consequence of the risk-aversion and the

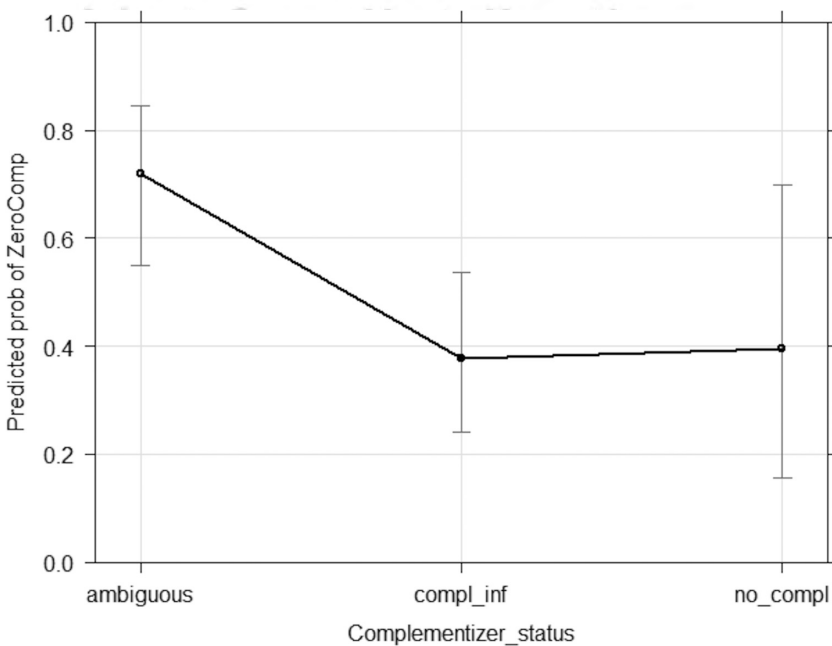


Figure 3.10 Effect plot of Complementiser Status in translated Dutch

processing-strain hypothesis. In our data, however, we do find a certain effect of source-language transfer, in that the explicit construction is relatively more frequent in Dutch texts translated from French than in Dutch texts translated from English, French being a language in which the presence of a complementiser relevantly similar to Dutch *om* in the IC is obligatory in virtually all cases whereas in English it is less clear whether infinitival *to* can be considered an explicit boundary signal on a par with Dutch *om*.

## 5. Conclusion

The aim of this chapter was to further disentangle the three proposed causes of explicitation in translations, following up on the studies by Kruger (2018) and Kruger & De Sutter (2018), which, on the basis of an investigation of the *that*/zero alternation in original and translated British and South African English, conclude that risk-aversion is the most plausible explanation, i.e. translators tend to opt for the safest option in an alternation in order to reduce the risk of communicative failure. However, in the case of the *that*-alternation investigated by Kruger & De Sutter (2018), the *that*-variant is the safest option in several different respects: it is the grammatically most explicit form, but at the same time it is the overall most frequent option and the choice most in line with perceived norms of good writing. By investigating an alternation in which the grammatically explicit option is not the overall most frequent one, we aimed to tease apart these different aspects.

The first step in our analysis of the *om*-alternation was to test whether the translators' choices significantly differ from the non-translators'. It turned out that translators more often retain the complementiser at the beginning of the IC while the default option in original Dutch is the implicit variant of the construction. By using the statistical MuPDAR procedure, we wanted to contribute to a better understanding of the factors underlying the difference between the two language varieties. First of all, the analyses showed that in both translated and non-translated Dutch the variation between the explicit and the implicit construction could be quite adequately modeled, with surprisingly few determining factors turning out as significant predictors. Register was found to have a significant effect on the choice in the original data set. The MuPDAR procedure indicated that the different behavior of translators and non-translators was to a large extent due to two predictors that differently affect grammatical choices in the two varieties, viz. the voice of the verb and the distance between the matrix clause head and the onset of the IC, both of which are complexity-related factors. With regard to both factors, it can be seen that translators tend to opt for the explicit construction more often than non-translators especially in contexts of low complexity. In a way, this can be seen as a risk-aversion strategy, which



would suggest that what is of the essence in the context of risk aversion is not so much frequency but degree of grammatical explicitness, i.e. the “safest” option in the *om*-alternation is the *om*-variant despite it being the overall least frequent option in non-translated language (though not by a very wide margin).

In the last part of Section 3, the third proposed hypothesis, the source-language transfer hypothesis, was examined through a third glmm based on the translated data, which showed that the choice between the explicit or implicit construction significantly differs according to the source language, next to other determining factors like the head of the IC, register and so on. For a better understanding of this difference, we took a closer look at the equivalents of the *om*-construction in the two source languages under scrutiny. It became clear that the structural “options” in French are obligatorily more explicit and that this preference is transferred to the Dutch target texts, even though Dutch tends to prefer a more implicit style. The situation in English was more complicated due to the more ambiguous status of *to* as an equivalent of rather *om* or *te*. Altogether, the data suggest that source-language transfer effects can indeed contribute to the greater preference for explicit structures in translated language.

In the future, we want to find a way to test the three hypotheses simultaneously in a model that also takes into account syntactic and semantic factors that were not yet investigated in the current study. Recently, De Sutter & Kruger (2018) have discovered some methodological issues in the MuPDAR approach after cross-validating the fitted model based on original English to an independent data set—again based on original English but extracted from another corpus (De Sutter & Lefer Under review)—and the factors that were significant in the first study turned out not to be in the second one. Moreover, a totally different set of factors was found to determine the choice in the second study, while the c-index and classification accuracy remained almost the same. These findings warrant, in other words, some caution for statistical analyses with MuPDAR.

For now, we hope to have shown that a possible scenario is for translators to prefer the more explicit construction even in cases where the implicit construction is actually the most common option in non-translated texts. In addition, it seems that the causes of explicitation can differ according to the phenomenon and the language pair(s) under investigation. While Kruger & De Sutter (2018) found evidence for both risk avoidance and, though to a lesser extent, processing effort, as the most important underlying mechanisms and ruled out a source language transfer effect (see, however, Footnote 2), the present study rather points in the direction of risk-aversion and source-language transfer as the most plausible causes. This implies that, in this stage, we cannot formulate a definitive answer to the question of why explicitation occurs in translated texts *in general*.

## Notes

1. All authors are affiliated with Ghent University. The first and second author are affiliated with the EQTIS research group, the first and third author with the GLIMS research group. The research reported here was enabled by a grant from the Ghent University Special Research Fund (BOF) for a project on “Grammatical alternations, register variation and explicitation in original and translated Dutch”. We would like to thank two anonymous reviewers for their helpful comments on an earlier version of the paper.
2. On a sidenote, however, it should be observed that, according to the corpus-based study by Colleman et al. (2016), the overall preference for the zero variant in present-day Afrikaans complement clauses is due to the lexical preferences of a small set of just three or four extremely high-frequency matrix verbs, whereas the large majority of Afrikaans matrix verbs actually prefer the explicit *dat*-variant (often by far). In this respect, Kruger’s (2018) observation that Afrikaans demonstrates a higher rate of complementizer omission than English and that, thus, under the source language hypothesis we would expect to see more *that*-omission in English translated from Afrikaans is perhaps in need of some qualification.
3. It should be noted that the presence of *om* is not always optional (Vandeweghe 2004: 329, 332–335; ANS 1997: 548–549, 557–559, 856–857, 899–900, 1110–1112; SoD 2015: 610–623). In contexts in which the IC functions as a purpose modifier clause as in *Ik heb het gedaan om hem te helpen* “I did it in order to help him” and in fixed expressions as *om het zacht uit te drukken* “to put it mildly” and *om kort te gaan* “in short”, *om* is obligatory. Conversely, the use of *om* at the beginning of an IC is impossible when the IC functions as the direct object of a number of verbs of saying, including *zeggen* “to say”, as in *Moeder zei (om\*) vroeg thuis te zullen zijn* “Mum said she would be home early”. Such cases were not included in the data set for the present study.
4. The DPC can be consulted via <http://dpcserv.ugent.be/comure> with login and password.
5. That is, in a context such as *Hij besloot om zijn job op te geven en (om) naar Spanje te verhuizen* “He decided to quit his job and move to Spain”, there is a tendency to leave out the second *om*.
6. The main analysis of Bouma in based on IC’s that depend on verbal heads. In a previous study, we also found evidence for the relevance of complexity-related factors, but in a different sample, viz. of IC’s in non-translated Dutch depending on *verbal* heads only (see Van Beveren et al. 2018).
7. In this extra test, we did not take into account explicit source texts with indicatives like *that*-clauses in English and *que*-clauses in French as in respectively “It is recommended that the system is primed before starting the infusion to the patient” and *le vice-président a suggéré qu’il vaudrait mieux renoncer à 25% de mon salaire*, because such sentences are not direct equivalents of the *om*-constructions in this case study.

## References

- ANS = Haeseryn, Walter, Romijn, Kirsten & Geerts, Guido. 1997. *Algemene Nederlandse Spraakkunst*. Groningen: Nijhoff and Deurne: Wolters Plantyn.
- Baayen, Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Baker, Mona. 1993. Corpus linguistics and translation studies: Implications and applications. In Baker, Mona, Francis, Gill & Tognini-Bonelli, Elena (eds.), *Text and Technology: In Honour of John Sinclair*. 233–250. Amsterdam: Benjamins.

- Becher, Viktor. 2010. Abandoning the notion of “translation-inherent” explicitation: Against a dogma of translation studies. *Across Languages and Cultures* 11. 1–28.
- Biber, Douglas. 2012. Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory* 8(1). 9–37.
- Biber, Douglas & Conrad, Susan. 2009. *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Blom, Alied. 1984. Enkele opmerkingen over *te* en *om*. *Spektator*. 170–176.
- Blom, Alied. 1990. *Om* voor de infinitief in *Eline Vere*: een weloverwogen keuze. *Spektator*. 9–18.
- Blum-Kulka, Shoshana. 1986. Shifts of cohesion and coherence in translation. In House, Juliane & Blum-Kulka, Shoshana (eds.), *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*, Vol. 272. 17–35. Tübingen: Gunter Narr Verlag.
- Bouma, Gosse. 2017. Om-omission. In Wieling, Martijn, Kroon, Martin, van Noord, Gertjan & Bouma, Gosse (eds.), *From Semantics to Dialectometry: Festschrift in Honor of John Nerbonne*. Tributes, Vol. 32. 65–73. UK: College Publications.
- Cappelle, Bert & Loock, Rudy. 2013. Is there interference of usage constraints? A frequency study of existential *there* is and its French equivalent *il y a* in translated vs. non-translated texts. *Target* 25(2). 252–275.
- Colleman, Timothy, Feinauer, Ilse & Braeckveldt, Charlotte. 2016. Over lexicale voorkeuren in de alternantie tussen de “skoon bysin” en de “dat-bysin”: een distinctieve collexeemanalyse. *Tydskrif vir Geesteswetenskappe* 56(1). 117–133.
- Cox, Thomas J. 1983. Teaching the unteachable: Prepositional complementizers in French. *The French Review* 57(2). 168–178.
- Delaere, Isabelle. 2015. *Do Translations Walk the Line?: Visually Exploring Translated and Non-Translated Texts in Search of Norm Conformity*. Doctoral dissertation. Ghent University, Ghent.
- De Sutter, Gert & Kruger, Haidee. 2018. Disentangling the motivations underlying syntactic explicitation in contact varieties: A MuPDAR analysis of that vs. zero complementation. (Presentation given at UCCTS 2018),
- De Sutter, Gert & Lefer, Marie-Aude. Under review. Empirical translation studies in the post-Baker era: A multifactorial, interdisciplinary and multi-methodological approach. *Perspectives: Studies in Translation Theory and Practice*.
- De Sutter, Gert, Speelman, Dirk & Geeraerts, Dirk. 2005. Regionale en stilistische effecten op de woordvolgorde in werkwoordelijke eindgroepen. *Nederlandse Taalkunde* 10(2). 97–128.
- Devos, Filip, De Muynck, Rik & Martens, Lutgart. 1992. *Nederlands, Frans en Engels in Contrast. 2: De zin*. Leuven: Peeters.
- Falk, Yehuda N. 2000. Infinitival to. (<http://pluto.huji.ac.il/msyfolk/Infinitival.pdf>).
- Grevisse, Maurice & Goosse, André. 2016. *Le Bon Usage: Grammaire Française*. Louvain-la-Neuve: De Boeck.
- Gries, Stefan & Bernaisch, Tobias. 2016. Exploring epicentres empirically: Focus on South Asian Englishes. *English World-Wide* 37(1). 1–25.
- Gries, Stefan & Deshors, Sandra. 2014. Using regressions to explore deviations between corpus data and a standard/target: Two suggestions. *Corpora* 9(1). 109–136.

- Hansen-Schirra, Silvia, Neumann, Stella & Steiner, Erich. 2007. Cohesive explicitness and explicitation in an English-German translation corpus. *Languages in Contrast* 7(2). 241–266.
- Hawkins, John. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- House, Juliane. 2011. Using translation and parallel text corpora to investigate the influence of global English on textual norms in other languages. In Kruger, Alet, Wallmach, Kim & Munday, Jeremy (eds.), *Corpus-Based Translation Studies*. 187–208. New York: Continuum.
- Huddleston, Rodney Desmond, Pullum, Geoffrey K. & Bauer, Laurie. 2006. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Jansen, Frank. 1987. Omtrent de om-trend. *Spektator* 17. 83–98.
- Klaudy, Kinga. 2004. The asymmetry hypothesis: Testing the asymmetric relationship between explicitations and implicitations. In Hansen, Gyde, Malmkjær, Kirsten & Gile, Daniel (eds.), *Claims, Changes and Challenges in Translation Studies: Selected Contributions from the EST Congress, Copenhagen*. Amsterdam: John Benjamins Publishing Company.
- Kolbe-Hanna, Daniela & Szmrecsanyi, Benedikt. 2015. Grammatical variation. In Biber, Douglas & Reppen, Randi (eds.), *The Cambridge Handbook of English Corpus Linguistics*. 161–179. Cambridge: Cambridge University Press.
- Kruger, Haidee. 2018. That again: A multivariate analysis of the factors conditioning syntactic explicitness in translated English. *Across Languages and Cultures* 20(1). 1–33.
- Kruger, Haidee & De Sutter, Gert. 2018. Alternations in contact and non-contact varieties: Reconceptualising that-omission in translated and non-translated English using the MuPDAR approach. *Translation, Cognition & Behavior* 1(2). 251–290.
- Kruger, Haidee & Van Rooy, Bertus. 2016. Syntactic and pragmatic transfer effects in reported-speech constructions in three contact varieties of English influenced by Afrikaans. *Language Sciences* 56. 118–131.
- Macken, Lieve, De Clercq, Orphée & Paulussen, Hans. 2011. Dutch parallel corpus: A balanced copyright-cleared parallel corpus. *Meta* 56(2). 374–390.
- Mondorf, Britta. 2014. (Apparently) competing motivations in morphosyntactic variation. In MacWhinney, Brian, Malchukov, Andrej & Moravcsik, Edith (eds.), *Competing Motivations in Grammar and Usage*. 209–228. Oxford: Oxford University Press.
- Olohan, Maeve. 2003. How frequent are the contractions?: A study of contracted forms in the Translational English corpus. *Target* 15(1). 59–89.
- Olohan, Maeve & Baker, Mona. 2000. Reporting that in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures* 1(2). 141–158.
- Øverås, Linn. 1998. In search of the third code: An investigation of norms in literary translation. *Meta* 43(4). 557–570.
- Pápai, Vilma. 2004. Explicitation: A universal of translation text? In Mauranen, Anna & Kujamäki, Pekka (eds.), *Translation universals: Do They Exist?*. 143–164. Amsterdam and Philadelphia: John Benjamins.
- Pardoen, Justine. 1998. *Interpretatiestructuur. Een onderzoek naar de relatie tussen woordvolgorde en zinsbetekenis in het Nederlands*, Uitgaven Stichting Neerlandistiek VU. Amsterdam/Philadelphia: Vrije Universiteit Amsterdam.

- Pullum, Geoffrey K. 1982. Syncategorematicity and English Infinitival to. *Glossa* 16. 181–215.
- Pym, Anthony. 2005. Explaining explicitation. In Karoly, Krisztina & Foris, Agota (eds.), *New Trends in Translation Studies: In Honour of Kinga Klaudy*. 29–34. Budapest: Akademiai Kiado.
- Pym, Anthony. 2015. Translating as risk management. *Journal of Pragmatics* 85. 67–80.
- Rohdenburg, Günter. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7(2). 149–182.
- Serbina, Tatiana. 2015. *A Construction Grammar Approach to the Analysis of Translation Shifts: A Corpus-Based Study*. Aachen: Rheinisch-Westfälischen Technischen Hochschule Aachen.
- SoD = Broekhuis, Hans, Corver, Norbert, Vos, Riet & Bennis, Hans. 2015. *Syntax of Dutch: Verbs and Verb Phrases*. Amsterdam: Amsterdam University Press.
- Stevenson, Angus. 2010. *Oxford Dictionary of English*. New York: Oxford University Press.
- Tagliamonte, Sali & Smith, Jennifer. 2005. No momentary fancy! The zero “complementizer” in English dialects. *English Language and Linguistics* 9(2). 289–309.
- Torres Cacoullos, Rena & Walker, James. 2009. On the persistence of grammar in discourse formulas: A variationist study of that. *Linguistics* 47(1). 1–43.
- Toury, Gideon. 2012. *Descriptive Translation Studies and Beyond: Revised Edition*. Amsterdam: John Benjamins.
- Van Beveren, Amélie, Coleman, Timothy & De Sutter, Gert. 2018. De om-alternantie. Een verkennende casestudy. *Handelingen: Koninklijke Zuid-Nederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis*. 191–222.
- Vandeweghe, Willy. 1972. Om en rond de (om) te-konstruktie. *Studia Germanica Gandensia*. 13. 37–61.
- Vandeweghe, Willy. 2004. *Grammatica van de Nederlandse zin*. Antwerpen and Apeldoorn: Garant.
- Van Haften, Ton. 1991. *De interpretatie van verzwegen subjecten*. Amsterdam: Vrije Universiteit Amsterdam.
- Vliegen, Maurice. 2001. Het facultatieve om na illocutionaire werkwoorden. *Nederlandse Taalkunde* 6. 112–132.